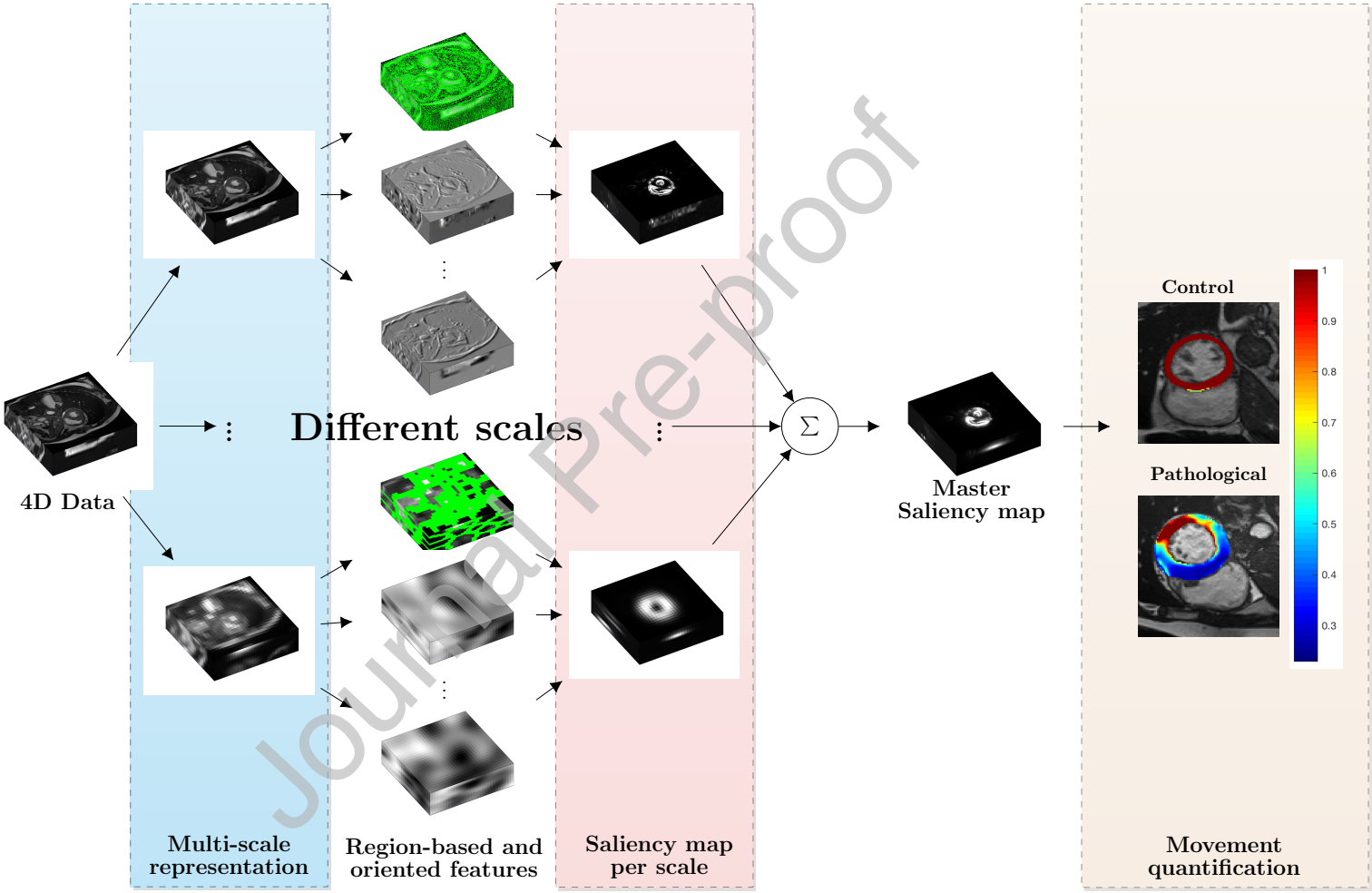
Characterization of motion patterns by a spatio-temporal saliency descriptor in cardiac cine MRI

Angélica Atehortúa, Eduardo Romero, Mireille Garreau

Highlights

- A 3D Spatio-temporal saliency descriptor generalizable to detect local cardiac motion
- Spatio-temporal saliency maps with clinical interest for the interpretability of the abnormal motion detection
- Spatio-temporal saliency method that correlates with abnormal patterns of hypertrophy cardiomyopathy at local and regional cardiac levels
- Characterization of motion patterns for early myocardial dysfunction, before development of the left ventricular hypertrophy

raphical Abstract



Characterization of motion patterns by a spatio-temporal saliency descriptor in cardiac cine MRI

Angélica Atehortúa[a,b], Eduardo Romero[a,*], Mireille Garreau[b]

[a] *Universidad Nacional de Colombia, Bogotá, Colombia*
[b] *Univ Rennes, Inserm, LTSI UMR 1099,F-35000 Rennes, France*

A B S T R A C T

Background and objective: Abnormalities of the heart motion reveal the presence of a disease. However, a quantitative interpretation of the motion is still a challenge due to the complex dynamics of the heart. This work proposes a quantitative characterization of regional cardiac motion patterns in cine magnetic resonance imaging (MRI) by a novel spatio-temporal saliency descriptor.
Method: The strategy starts by dividing the cardiac sequence into a progression of scales which are in due turn mapped to a feature space of regional orientation changes, mimicking the multi-resolution decomposition of oriented primitive changes of visual systems. These changes are estimated as the difference between a particular time and the rest of the sequence. This decomposition is then temporarily and regionally integrated for a particular orientation and then for the set of different orientations. A final spatio-temporal 4D saliency map is obtained as the summation of the previously integrated information for the available scales. The saliency dispersion of this map was computed in standard cardiac locations as a measure of the regional motion pattern and was applied to discriminate control and hypertrophic cardiomyopathy (HCM) subjects during the diastolic phase.
Results: Salient motion patterns were estimated from an experimental set, which consisted of 3D sequences acquired by MRI from 108 subjects (33 control, 35 HCM, 20 dilated cardiomyopathy (DCM), and 20 myocardial infarction (MINF) from heterogeneous datasets). HCM and control subjects were classified by an SVM that learned the salient motion patterns estimated from the presented strategy, by achieving a 94% AUC. In addition, statistical differences (test t-student, $p < 0.05$) were found among groups of disease in the septal and anterior ventricular segments at both the ED and ES, with salient motion characteristics aligned with existing knowledge on the diseases.
Conclusions: Regional wall motion abnormality in the apical, anterior, basal, and inferior segments was associated with the saliency dispersion in HCM, DCM, and MINF compared to healthy controls during the systolic and diastolic phases. This saliency analysis may be used to detect subtle changes in heart function.

Keywords: Heart motion, cardiac cine magnetic resonance imaging, spa-tio-temporal saliency, cardiac motion patterns

1. Introduction

1.1. Clinical context and motivation

Characterization of heart diseases is not an easy task because of the several compensatory mechanisms that may hinder the effect exclusively caused by the disease (Mann and Bristow (2005)). Pathologies like hypertrophic cardiomyopathy (HCM) are asymptomatic and present an impaired regional myocardial function at the diastole and systole, with preserved or slightly increased ejection fraction (Marian and Braunwald (2017); Weissler-Snir et al. (2017)). This scenario of few or no changes at all complicates treatment personalization, above all if one considers how the cardiac structure is transforming to maintain the hemodynamic physiology (Elliott et al. (2014); Weissler-Snir et al. (2017)). In this context, new cardiac characterization strategies are required to understand and individualize disease adaptation mechanisms (Editorial (2018)). This disease knowledge may facilitate an earlier HCM detection, being useful to reduce the risk of sudden cardiac death in about 25% of HCM patients (Marian and Braunwald (2017); Weissler-Snir et al. (2017)), as well as, to improve the poor prognosis in other complex and asymptomatic diseases.

By directly analyzing the temporal dynamics of the heart, subtle and compensatory changes during diastolic and/or systolic functions may be detected (Mazein et al. (2018); Buckberg et al. (2008); Omar et al. (2015)). They may reveal abnormal movement patterns which can be associated with actual disease mechanisms. In this context, quantitative temporal measurements can follow the cardiac function, despite the variable acquisition protocols, pathology conditions, and complex physiology. However, these challenging scenarios require analysis methods with high resilience to different types of noise for extracting relevant information.

1.2. Technical context

Recent machine learning strategies have accurately predicted the patient's risk of heart failure by tracking the motion in cine MRI (Bello et al. (2019); Krittanawong et al. (2019); Leiner et al.; Sarmiento et al. (2018)). Deep learning models have been used to predict outcomes from dense phenotypic data in heart time-resolved segmentation (Bello et al. (2019)). In these works, the motion has been tracked using a B-spline image registration between consecutive 3D heart segmentations to estimate a motion field, which is represented by the displacement vector at each voxel and at each instant of the cardiac cycle. Despite these results, deep learning strategies are dependent on the dataset size, the image quality, the available annotations, and the standardization. Unlike these approaches, the strategy herein presented, based on the human visual attention model, extracts relevant information in complex scenarios with less dependence on the number or quality of data, as demonstrated by proving the robustness of this method to different types of noise, namely Rician, speckle and Gaussian and different configurations of the Gaussian noise. Likewise, the method performed well when classifying different databases, showing relative independence of the number of cases to train.

Visual attention models have been used to extract meaningful image content in very noisy conditions. Current computational models of saliency, inspired by the visual system, select information by applying a bottom-up scheme, i.e, a location is highlighted if it is different from its surroundings and similar to other regions in the image. These saliency models have been broadly applied to medical images in problems related with anomaly detection (Zhao et al. (2017); Dubost et al. (2019)), region of interest detection (Iakovidis et al. (2018); Huang et al. (2018)), registration (Nguyen and Wu (2016)) and segmentation (Mahapatra and Buhmann (2016)). Lately, deep learning strategies have been introduced to highlight salient regions in the brain, breast, and dermoscopy, images (Zhang et al. (2019, 2018); Ding et al. (2019)) and in the heart to detect artifacts of capture (Oksuz et al. (2019)). All these approaches exclusively lie in spatial information. The temporal component of the cardiac function is a crucial feature not only to find out hidden characteristic movement patterns but to interpret the disease mechanisms.

1.3. Proposed approach

In this paper, a new multi-scale region-based spatio-temporal saliency map descriptor is presented, named as STSalmap (Spatio-Temporal SALiency MAP). This descriptor is inspired by the visual attention system, aiming to extract relevant temporal information in any dynamic scene. The main contributions of this work are summarized as follows:

- The formulation of a novel spatio-temporal saliency description set upon a set of simple techniques which requires little parameter tuning and data for training, i.e., multi-resolution analysis by oriented Gaussian kernels that integrates temporal information and 3D multiscale representations, making this model more robust to capturing noise, variability of the vendor and the specific pathology.
- A spatio-temporal saliency framework devised to characterize subtle changes of complex cardiac pathologies, capable of distinguishing between healthy and HCM subjects, with comparatively lower false-positive rates than other tested methods.
- An interpretable indicator that provides summarized information of the cardiac function that can complement the traditional cardiac indexes like the ejection fraction, which has been reported to show normal values in some cases of heart failure or hypertrophic cardiomyopathy cases.
- A model evaluated against several datasets, specifically four different collections of cases, demonstrating the generalizability of the proposed method.

* Corresponding author.
e-mail: edromero@unal.edu.co (Eduardo Romero)

This paper is organized as follows: a background on spatio-temporal saliency methods is presented in Section 2. Section 3 introduces the motion perception model, which has inspired the proposed spatio-temporal saliency descriptor. Section 4 introduces the spatio-temporal descriptor. Experiments and results are provided in Section 5. Finally, discussion and conclusion about the proposed strategy are presented in Sections 6 and 7 respectively.

2. Related spatio-temporal saliency methods

This section presents different spatio-temporal saliency methods in video applications. Few methods have estimated saliency in video sequences, organized into three groups: i) approaches using optical flow to extract motion information, ii) methods based on learning a saliency model, and iii) other strategies.

i) Optical flow approaches. The resulting flow, combined with low-level features such as edges, intensity and color, serves to regionally perform intra-frame differences. These approaches, performed by regions known as superpixels, obtain a saliency as a probability map of the intra-frame difference, modulated by background information (Wang et al. (2018); Chen et al. (2018); Liu et al. (2014, 2017); Wang et al. (2015)). The main limitation of these strategies is their dependence on the regularization equation that governs the amount of allowed motion to track and when an object location changes abruptly, optical flow algorithms would generally not perform so well.

ii) Deep learning to model saliency. Tang et al. (2018) trained a multi-scale spatio-temporal convolutional network to predict saliency by mixing coarse labels with fine labels of saliency information. Bak et al. (2018) built a spatio-temporal net by learning spatial correlations between adjacent frames, while Jiang et al. (2018) used a convolutional neural network to predict intra-frame saliency, using a kind of objectness estimation and motion subnets. A great limitation with these strategies is that they require lots of manually annotated data, a real problem when dealing with cardiac MRI. In addition, these methods tend to overfit and overparameterize the classification model, leading to false-positive results and also overestimating the image saliency. Other recent state-of-the-art methods (Sun et al. (2018); Faisal et al. (2020); Patil et al. (2019); Maczyta et al. (2019)) use convolutional neural networks to learn the motion saliency in nature images yet. These methods use a lot of training data to learn its variability and combine other classical strategies as optical flow to extract relevant information. Data availability is a major limitation for applying these methods in medical images, which in addition are much more complex than natural images.

iii) Other strategies. Kim et al. (2015) modeled motion as a Markov process in natural images, using a graph with transition probabilities among the graph nodes and each node being a state of the process. A temporal saliency distribution is obtained as a steady-state distribution of a random walker running over such a graph. Spatial features of intensity, color, and compactness are used to estimate the transition between states. However, this strategy is based on assumptions that are impossible to fulfill in data with high variability, absence of color, and subtle differences between objects.

The proposed method uses neither optical flow nor saliency learning, but it integrates temporal information, unlike most saliency methods in medical imaging. Spatio-temporal saliency models may simply correlate subtle morphological abnormalities during the cardiac cycle by capturing the variability of motion of these regions. In contrast to the state-of-the-art spatio-temporal saliency methods, designed for 2D natural images, the proposed descriptor uses the third dimension and obtains a better representation of the heart movement. Given the 3D shape nature of the cardiac structures, global and compact information may be estimated even in fuzzy slices like the apex. In addition, the proposed descriptor follows a motion perception model, and therefore the spatio-temporal salient information is captured even in noisy environments. Likewise, a minimal number of the model parameters has to be tuned, and then no training or huge quantities of data are needed to capture the high structure and dynamics variability of the heart. This motion perception model is described in the following section 3.

3. Introducing the motion perception model

The visual system reacts to changes or discontinuities and captures the salient spatio-temporal information by estimating several features or primitives with coherent motion during a sequence, even in noisy conditions (Johansson (1973)). Motion processing is carried out at two different levels: the primary visual cortex and the middle temporal (MT) area (Treue and Maunsell (1996); Albright and Stoner (1995)). At the former, the first processing consists in locally detecting oriented spatio-temporal discontinuities of low-level primitives, namely contrast, texture, and edges. Relevant primitives are small regions moving with different orientations and velocities. A global velocity estimation is then constructed at the latter area (Albright and Stoner (1995)), the MT area, which integrates the output of the first area and defines a preferred direction of motion which serves as a global reference of the moving primitives. Different primitives may be coherently perceived from the built-up reference (Schrater et al. (2000)).

The proposed STSalmap descriptor aims to mimic this dynamic by estimating the coherent motion of relevant features in a sequence. For doing so, every sequence frame is firstly decomposed in multiple scales which are divided into small irregular regions, each containing oriented responses from low-level features. The motion of these irregular regions is estimated at any time t by superimposing the partition at time t to the rest of the sequence. Afterward, the sum of the squared

differences between the frame at time t and the other frames is computed at each region, obtaining an oriented regional motion. Since the energy of these motions are uniformly distributed over all orientations, these oriented motions are then integrated, approximating a regional speed pattern, i.e., the local motion corresponds to the spatio-temporal locations where energy is concentrated (Schrater et al. (2000); Adelson and Bergen (1985)). Finally, these results are averaged in the distinct scales to remove uncorrelated noise and preserve the most relevant information, even in short or large displacements. This information represented the coherent or global motion of moving features at any instant of time t of the sequence, being it correlated with psychological studies reported by Schrater et al. (2000) and Bradley and Goyal (2008). This global motion was here defined as the motion saliency map.

4. Method

4.1. Overview

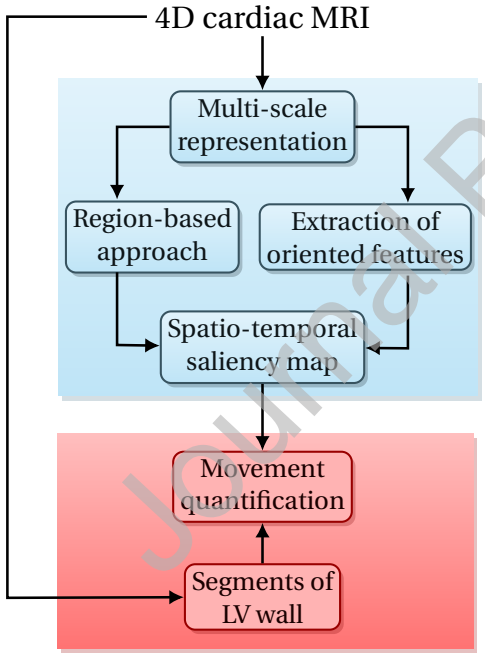


Fig. 1: Flowchart of the proposed method: The STSalmap descriptor (blue area) estimates the motion saliency from 4D data. Then, movement patterns are characterized from this saliency (red area), in anatomical segments of the left ventricle.

This section briefly introduces a spatio-temporal saliency descriptor to characterize cardiac movement patterns. The workflow is shown in Fig. 1. The proposed STSalmap descriptor starts by mapping the original image to a multi-scale space, as described in subsection 4.2.2, to generate a per-scale-coarse spatial saliency (subsection 4.2.3). Likewise, a region-based approach splits each of the different scales into a set of disjoint regions, as presented in subsection 4.2.4. Afterward, a spatio-temporal saliency map is regionally estimated as an average of the saliency at each of the scales, as further explained in subsection 4.2.5.

Finally, regional left ventricular (LV) wall segmentations at end of the diastole and the end of the systole are superimposed to the obtained saliency and movement patterns are characterized by computing a dispersion measure of the saliency at each of the myocardium segments.

4.2. Spatio-temporal saliency map descriptor (STSalmap)

4.2.1. Problem formulation

Let us denote an image $\mathrm{I} \subset \mathscr{R}^3$, being (x, y, z) the spatial coordinates of I. A temporal sequence is defined as the set T of times t, $\mathrm{T} = \{t = 1, t = 2, \ldots, t = \mathrm{N}\}$, where N is the number of frames. Hence, an image at time $t \in \mathrm{T}$ is defined as $\mathrm{I}(t)$ and the STSalmap descriptor maps $\mathrm{I}(t)$ to its corresponding salient image $\mathrm{S}(t)$ by quantifying regional spatio-temporal changes of oriented features during the sequence T.

4.2.2. Multi-scale representation

A multi-scale analysis is applied to extract the most relevant information preserved through the scales. Each image $\mathrm{I}(t)$ from the temporal sequence T is decomposed into several s scales, $\mathrm{I}(t, s)$, by applying a Gaussian pyramidal decomposition (Burt and Adelson (1987)). This pyramid is constructed by iteratively smoothing and down-sampling/up-sampling (reduction/expansion direction) the image. Smoothing is carried out for each voxel of the image as a weighted average of its neighboring voxels. The weights are computed by the kernel w, defined as $w = [\frac{1}{4} - \frac{a}{2}, \frac{1}{4}, a, \frac{1}{4}, \frac{1}{4} - \frac{a}{2}]$ (Burt and Adelson (1983)), being a the parameter that controls the shape of this kernel. In this work, $a = 0.375$, for which w is approximated to a Gaussian distribution. Then, this smoothed image is dyadically down-sampled/up-sampled along each direction, without modifying the length of the third dimension of the image. This scale factor has been commonly used in image analysis trying to remove non-correlated noise and to preserve the scale invariance law (Florack et al. (1992)). In this work, the multi-scale representation is set to three coarser levels, the original scale, and a single finer level.

4.2.3. Extraction of oriented features

Visual perception of moving structures is a quite robust mechanism even under very noisy conditions, e.g., fuzzy edges, superposition, occlusion, among others. In particular, most imaging techniques of the cardiac dynamics are contaminated by different types of noise, being inherent to the properties of each physical principle of the image. Gaussian, Poisson, and Rician correspond to typical noises found in medical imaging. Movement patterns are then difficult to characterize, especially for cardiac problems because the heart seems to move slower or less than it moves (Massanes and Brankov (2011)). The main purpose of this investigation is to find out relevant moving features and their basic relationships. The first step is to convert low-level primitives of the image, such as orientation, contrast, and intensities, into

relevant information by applying an oriented filter (Freeman and Adelson (1991)), which mimics the basic directional decomposition achieved by the retina neurons in the first phase of the perception phenomenon (Mather et al. (2013)). This filter is applied to every image $I(t,s)$ at time t of the temporal sequence T and at each of the scales s, obtaining a salient orientation $F(t,s,\theta)$ at arbitrary θ direction. This oriented filter is an orientation-selective convolution kernel G of the oriented first derivative of a 2D Gaussian function $f(x,y)$, being the 2D Gaussian function defined as:

$$f(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(\frac{-(x^2+y^2)}{2\sigma^2}\right) \tag{1}$$

Where the first x derivative of this Gaussian $f(x,y)$ rotated by an angle 0° about the origin defined as

$$G^{0^\circ} = \frac{\partial f(x,y)}{\partial x} = -\frac{x}{\sigma^2}\frac{\exp(\frac{-(x^2+y^2)}{2\sigma^2})}{\sigma\sqrt{2\pi}} \tag{2}$$

And the first y derivative of that same function, rotated 90°, corresponds to

$$G^{90^\circ} = \frac{\partial f(x,y)}{\partial y} = -\frac{y}{\sigma^2}\frac{\exp(\frac{-(x^2+y^2)}{2\sigma^2})}{\sigma\sqrt{2\pi}} \tag{3}$$

Being G^0 and G^{90} the basis filter. Since convolution ($\circledast$) is a linear operator, the image $I(t,s)$ can be synthesized at an arbitrary orientation by performing linear combinations of the image $I(t,s)$ with the basis filter as follows:

$$R^{0^\circ} = G^{0^\circ} \circledast I(t,s) \tag{4}$$

$$R^{90^\circ} = G^{90^\circ} \circledast I(u,t,s) \tag{5}$$

The oriented filter (G^{0° and G^{90°) is applied to the 3D data by moving the first derivative Gaussian 2D kernel slice per slice. Finally, the oriented response in projection functions $\cos\theta$ and $\sin\theta$ of the features in the image $I(t,s)$, at the arbitrary θ direction is

$$F(t,s,\theta) = \cos(\theta)R^{0^\circ} + \sin(\theta)R^{90^\circ} \tag{6}$$

The intensity of the original input image is filtered by the orientation filter whose output $F(t,s,\theta)$ is normalized per volume by using min-max algorithm. Notice how this approximation corresponds to an actual continuous representation in discrete terms. In this work, following orientations were used Θ ={0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°}. The σ parameter of this filter (Eq 1) was set to 3 after an exhaustive exploration, described in Section results 5. The same σ value was used in all scales in the multi-scale approach of the STSalmap descriptor since this parameter is invariant scale. Unlike Gabor filters, widely used, these oriented filters are computationally efficient since any arbitrary orientation can be obtained as a linear combination of a few filter responses (Huang and Chen (1995)).

4.2.4. Region-based approach

Motion is estimated by computing regional spatio-temporal changes of low level features at arbitrary directions. Localized changes are herein estimated by quantifying the L2-norm of pairwise spatio-temporal differences of oriented features, computed for each region of the scene. This regional information correlates with objects moving during a sequence and is computed within regions defined by a supervoxel partition (SLIC) (Achanta et al. (2012)) which matches region boundaries (supervoxels) with edges of structures. For doing so, SLIC algorithm uses a 4-dimensional (4D) space to represent a region $I(t,s)$ as (l,x',y',z'), being l the intensity and (x',y',z') the region spatial position. Given L, a desired number of regions, the image $I(t,s)$ is spatially divided in a grid of L similarly-sized regions, each of them with a center (l_i,x'_i,y'_i,z'_i), and $i = 1,\ldots,L$. Every region is then iteratively modified by assigning a voxel to the closest region center and updating the region center location. At each iteration, the gradient of every $3\times3\times3$ neighborhood is computed and the center of each region is set to the spatial position of the smallest gradient, avoiding the center coinciding with an edge. The closest region center i for each voxel j is then determined by a metrics D that combines intensity and spatial distances as follows:

$$d_{intensity} = \sqrt{(l_j - l_i)^2} \tag{7}$$

$$d_{spatial} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2} \tag{8}$$

$$D = \sqrt{\left(\frac{d_{intensity}}{m}\right)^2 + \left(\frac{d_{spatial}}{Q}\right)^2} \tag{9}$$

where Q is the initial distance between two adjacent region centers and m is a compactness parameter, a scalar value that controls the shape of the region using both intensity similarity and spatial proximity. This parameter is herein regionally adapted by forcing the region center to be the mean of all voxels belonging to the supervoxel, up to 10 iterations, since this number has demonstrated to be enough to find regions with minimum intra-region variance (Achanta et al. (2012)). However, the SLIC method may mismatch region boundaries and structures by an unsuitable number of regions L. Hence, this L parameter is adaptively estimated for each image $I(t,s)$ aiming to match an initial regular tessellation with regions (supervoxels) of size $\omega\times\omega\times\omega$. In this paper, $\omega = 7$ is set as suggested by Kervrann et al. Kervrann and Boulanger (2006). The number of regions $L(t,s)$ at scale s and time t is here then defined as:

$$L(t,s) = Round\left(\prod\frac{Dim(I(t,s))}{\omega} + median(Dim(I(t,s)))\right) \tag{10}$$

being $Dim(I(t,s))$ a vector whose elements are the corresponding spatial dimensions of the image $I(t,s)$ and Round the $L(t,s)$ value to the nearest integer. Once this L number of regions is obtained, the supervoxel partition is performed for each scale of the sequence $I(t,s)$, as previously mentioned,

to deform the initial tessellation to irregular regions whose edges remain well aligned with structural boundaries. Each obtained region $r(t,s)$, $r \in \mathrm{R}^3$, from the image $\mathrm{I}(t,s)$ represents the spatial coordinates of the supervoxel, contained in the spatial domain of the image at time t and scale s. The set of regions for each image is then stored in a grid $\mathrm{L_G}(t,s)$ to be used in the next step.

4.2.5. Spatio-temporal saliency map

The non-linear dynamic of the cardiac motion describes a series of smooth and abrupt changes during the same cardiac cycle, different at global and local levels, i.e., parts of the heart move more, and others remain static during different phases of the cycle. The proposed motion saliency model captures local and global changes of the heart during the cardiac sequence as follows.

Local salient motion.. The local saliency of a region $r \in \mathrm{I}(t,s)$ is estimated along the sequence of images $\mathrm{I}(\mathrm{T},s)$ by locally detecting oriented spatio-temporal discontinues in each direction θ. For doing so, $\mathrm{I}(\mathrm{T},s)$ is filtered as described in subsection 4.2.3, obtaining a set of oriented features $\mathrm{F}(\mathrm{T},s,\theta)$. The grid $\mathrm{L_G}(t,s)$ at time t and scale s is superimposed to the whole sequence $\mathrm{F}(\mathrm{T},s,\theta)$, and using this grid as spatial support, the differences between the features $\mathrm{F}(t,s,\theta)$ at time t and the rest of the sequence frames T is computed. Finally these differences are summed up to obtain:

$$\mathscr{E}(r,t,s,\theta) = \sum_{\tau=1}^{\mathrm{N}} \|\mathrm{F}(r,t,s,\theta) - \mathrm{F}(r,\tau,s,\theta)\|^2 \qquad (11)$$

where N is the number of images of the sequence. Equation 11 describes the summation of the spatio-temporal changes (map of changes) with respect to the time t of the sequence. This regional oriented motion (Eq 11), after integration along the different directions θ, yields the regional speed pattern of the salient features $\mathrm{S}(r,t,s)$, at time t and scale s:

$$\mathrm{S}(r,t,s) = \sum_{\theta\in\Theta} \mathscr{E}(r,t,s,\theta) \qquad (12)$$

Global salient motion. The regional speed pattern is simply lower when the region visual information is uniformly distributed and much higher when it is concentrated in specific orientation bands, imitating thereby what has been reported in psychophysical experiments of the human visual system (Schrater et al. (2000)). This speed pattern is computed for every region $r \in \mathrm{L_G}(t,s)$, obtaining a saliency map $\mathrm{S}(t,s)$ for each scale. Afterward, each of these scaled saliency maps is interpolated $\hat{\mathrm{S}}(t,s)$ to the spatial dimensions of the original image, using the nearest-neighbor algorithm, and averaged to obtain the global motion saliency, i.e., a master spatio-temporal saliency map $\mathrm{S}(t)$:

$$\mathrm{S}(t) = \frac{1}{5}\sum_{s=1}^{5} \hat{\mathrm{S}}(t,s) \qquad (13)$$

This average is a low-pass filter that further removes uncorrelated noise and preserves the most relevant information in all scales(Allard and Arleo (2017)).

4.3. Characterization of movement patterns in cine MRI

Abnormal movements of segments of the left ventricle (LV) have been reported during the cardiac cycle in cases of cardiac failure (Rickers et al. (2005); Mishiro et al. (1999); Sanchez-Martinez et al. (2018); Garg et al. (2017)). For this reason, motion characterization of the left ventricle is herein carried out at the level of LV segments, as follows.

4.3.1. Anatomical segments of the left ventricular wall

Currently, the LV function is more precisely estimated by a regional examination of a set of anatomic segments of the LV wall, as recommended by the American Heart Association (Cerqueira et al. (2002)). Usually, this analysis is manually carried out per individual, a delicate time-consuming task that may introduce error by the high intra-and inter-variability heart shapes. An automatic LV wall segmentation is herein proposed, in which the wall is segmented into four segments: the first three, septal, anterior, and inferior are those parts nearby the septum, lung, and liver respectively, and the fourth lateral segment is the free wall (as observed in Fig. 2). Likewise, in the plane orthogonal to the cardiac long axis, the LV is divided into four circular sections: basal, mid-cavity, apical, and the apex section at the end of the ventricular chamber, composed of those slices not showing any LV cavity at all.

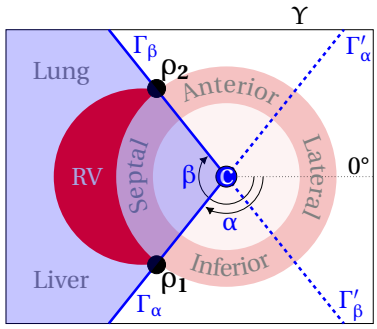


Fig. 2: LV wall is divided by the lines Γ. RV and LV cavities and LV myocardium are illustrated in red, light pink, and dark pink, respectively.

Figure 2 illustrates how the LV wall in the short axis slice Υ is automatically parcelled, as long as both right (RV) and left ventricles are segmented. For each slice, the strategy consists in tracing two lines (in blue), which must pass by the LV centroid $c = (x_c, y_c)$ and connect opposite sides of the border of the slice Υ. These two lines, Γ_α and Γ_β in Fig. 2, must in addition cross the closest point of the external surface of the right ventricle (ρ_1 and ρ_2 in Fig. 2). An exhaustive search at every clockwise degree sets the two lines defined by $x_j = j\,cos(\lambda) + x_c$ and $y_j = j\,sin(\lambda) + y_c$ with $j = 0,1,2,\ldots,k$, being k the border partition in degrees and λ a particular direction. As shown in Fig. 2 these two lines define the different wall segments, i.e., septal, anterior, lateral, and inferior. Likewise, the basal, mid, and apical sections along the long axis are obtained by simply dividing the LV cavity into thirds, while the fourth section, the apex, corresponds to those slices without any LV cavity.

4.3.2. Movement quantification in cine MRI

Once the LV myocardium segments are obtained, salient information may be quantified. Intuitively, high values of saliency are associated with more motion, i.e., the more a segment constantly moves, the greater its saliency is. This analysis can be useful in specific disorders like hypertrophic cardiomyopathy, where the non-homogeneous muscle growth may be correlated with a scattered saliency by the presence of both rapid and slow-motion patterns of the whole myocardial wall during the cardiac cycle (Kramer et al. (1994); Marian and Braunwald (2017)).

In this context, a measure of the saliency dispersion, the coefficient of variation CV, quantifies the segment movement. The higher this value, the greater the motion variability. This coefficient is dimensionless and scale-invariant, useful to describe data variation with respect to its mean, i.e., this approach facilitates comparison between different subjects by removing the influence of the heart size. This coefficient is defined as:

$$\mathrm{CV}_{seg} = \frac{\sigma_{seg}}{\mu_{seg}}, \qquad seg = 1, \ldots, \mathrm{N}_{seg}$$

being σ_{seg} and μ_{segm} the standard deviation and mean of the saliency computed in the segment seg. This coefficient is computed for all segments, N_{seg}, in which the left ventricular wall was divided.

5. Results

5.1. Data

The STSalmap descriptor was applied to short-axis cine MRI images from four public available cardiac image datasets. These datasets are:

Automated Cardiac Diagnosis Challenge, ***ACDC*** *(Bernard et al. (2018)).* Eighty cine MRI cases were used from this database. Images were acquired, over six years, with two MRI scanners of 1.5 T (Siemens Area) and 3.0 T (Siemens Trio Tim) (Siemens Medical Solutions), with spatial and temporal resolutions of $1.5 \pm 0.2\ mm^2/pixel$ and 27.1 ± 6.1 frames per cycle, respectively and a slice thickness of $9.3 \pm 1.7\ mm$ (Bernard et al. (2018)). These cases were divided into four subgroups of 20 individuals:

- Control.
- Previous myocardial infarction (MINF) with and an ejection fraction of the LV (LVEF) lower than 40% and myocardial segments with abnormal contraction.
- Hypertrophic cardiomyopathy (HCM) with an LV cardiac mass higher than $110\,g/m^2$, several myocardial segments with a thickness higher than $15\ mm$ in diastole and a normal LVEF
- Dilated cardiomyopathy (DCM) with diastolic left ventricular volume > $100\ m\mathrm{L}/m^2$ and an LVEF lower than 40%.

Cardiac MR Left Ventricle Segmentation Challenge MICCAI 2009, ***MI09*** *(Radau et al. (2009)).* Twelve cases were acquired with an MRI scanner of 1.5 T GE Signa MR at spatial resolution and slice thickness of $1.4 \pm 0.1\ mm^2/pixel$ and $8.9 \pm 1.0\ mm$ respectively, and temporal resolution of 20 cardiac frames over the heart cycle. The used data comprises two subgroups: three control cases with no hypertrophy but LVEF > 55% and nine HCM cases with normal LVEF and a ratio of LV mass over body surface area higher than $83\,g/m^2$.

Cardiac MRI dataset from York University, ***YORK*** *(Andreopoulos and Tsotsos (2008)).* Seven cases (one control and six HCM) were acquired with an MRI scanner of 1.5 T GE Genesis Signa MR at spatial resolution and slice thickness of $1.4 \pm 0.2\ mm^2$ and $8.9 \pm 1.5\ mm$, respectively, a temporal resolution of 20 frames per cardiac cycle. All the cases were under the age of 18.

Motion tracking challenge from STACOM 2011, ***MI11*** *(Tobon-Gomez et al. (2013)).* Nine control cases (aged 28 ± 5 years) were used from this dataset. Images were acquired with an MRI scan of 3 T Philips Achieva System, with spatial and temporal resolutions of $1.2 \pm 0.1\ mm^2\ pixel$ and 23.4 ± 8.3 frames per cycle, respectively, and a slice thickness of $8\ mm$.

Each dataset provides manual delineations of the left ventricle (LV) myocardium and right ventricle cavity (RV) at the end of the diastole (ED) and the end of the systole (ES).

5.2. Evaluation

The present method associates the characterized movement patterns with control and pathological subjects from heterogeneous datasets of cine MRI images. Such association was twofold demonstrated: the former consisted in discriminating control and hypertrophic (HCM) subjects with an SVM (Gaussian Radial Basis Function) classifier (Subsection 5.4) and the latter, in finding out significant differences of regional movement patterns between control and subjects diagnosed with HCM, dilated cardiomyopathy and myocardial infarction (Subsection 5.5).

5.3. Setting parameters

Only two parameters required some tuning. Firstly, the standard deviation σ of the oriented filter (Equations 2 and 3) which was set to 3 after exhaustive exploration ($\sigma = \{0.2, 0.5, 3, 5, 8, 10\}$). Secondly, since the image is divided into a number of regions L in the region-approach step (Eq. 10 of subsection 4.2.4), such number was computed by the adaptive strategy already described in the same section, with initial region size of $\omega = 7$ after exploring with $\omega = \{7, 11\}$ and number of regions set to $\mathrm{L} = \{100, 1000, 2000\}$.

5.4. Movement patterns of control and HCM subjects

This part of the evaluation was addressed to show that the motion characterization separates control and HCM patients in a classification task.

5.4.1. Experimental setup

Features. For each subject, saliency was computed at the ED using only volumes from the diastolic phase to evaluate the diastolic function as an indicator of disease. The left ventricle (LV), at the ED, was partitioned into four anatomical segments, as explained in subsection 4.3.1. A dispersion saliency measure was computed in each anatomical segment at the ED (movement patterns as described in subsection 4.3.2). As result, each subject was characterized by four features representing the saliency dispersion at different segments in the diastolic phase.

Classification. The experimental group from the whole set of heterogeneous databases consisted of 33 control subjects and 35 HCM subjects. A support vector machine (SVM) classifier with radial basis function (RBF) kernel was used to separate control and HCM subjects. Parameters of this kernel were automatically determined by a grid search with 3-fold cross-validation. A Z-score normalization was applied to features of each training and testing data. The classification was performed in two experiments:

Experiment 1: The SVM classifier was trained and tested using several K-fold cross-validation setups, with K = $\{3,5,7\}$ and using all datasets.

Experiment 2: The SVM classifier was trained only with the ACDC dataset (40 subjects, 20 control, and 20 HCM), and the other datasets were used for testing.

Performance evaluation. Performance was quantified by different metrics, namely accuracy, sensitivity, specificity, precision, area under the curve (AUC) and F-measure. F-measure was defined as $2 * ((\text{precision} \times \text{recall})/(\text{precision} + \text{recall}))$;.

Baselines. Since the proposed strategy estimates a spatio-temporal saliency, the same protocol of experimentation was applied to the methods presented by Liu et al. (2014) (Liu), Wang et al. (2015) (Wang) and Kim et al. (2015) (Kim). These three methods were here selected for two reasons: they estimate spatio-temporal saliency and their source codes were available. The first two approaches used optical flow to estimate saliency, while Kim et al. modeled motion as a Markov process.

Figure 3 illustrates the saliency obtained at the ED in the basal slice from a control subject by several spatio-temporal saliency methods. Note how Liu and Wang methods highlighted the ventricular cavities as the more salient structures, but practically no local motion information is observed. Some regional motion is captured by both Kim's and the proposed method. Nevertheless, the structures highlighted by Kim are blurry and practically non-interpretable. In contrast, the present method nicely delineates the cardiac structures and it is easy to observe, as expected, how the motion pattern is localized in the free wall of the left ventricle and lateral wall of the right ventricle.

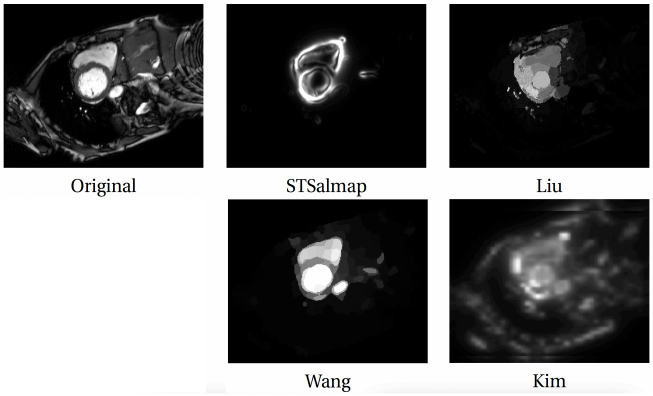


Fig. 3: Spatio-temporal saliency methods. The original image corresponds to the short axis MRI from a control subject at the ED in the basal slice.

5.4.2. Classification results

Experiment 1: K-fold cross-validation setups. Three setups, with K = $\{3,5,7\}$, were used to assess the aptness of this characterization to separate control subjects from HCM subjects. Results for each K-fold are shown in Table 1. Note how the STSalmap descriptor outperformed baseline methods for each of the metrics and most K-fold groups. Kim obtained similar F-measure values (in about 0.71), but overall their other metrics show lower values with higher standard deviations. Liu reported an average F-measure of 0.44 and Wang's F-measure values were not included since subjects were all classified as HCM in all folds. These results look consistent with what has been shown in Fig. 3, i.e., Liu's estimated motion saliency looks parceled and with no relation with ventricular structures while the other methods are concentrated within the cardiac cavities but Kim's method also highlights motion from breathing movements within the thoracic box. In addition, significant statistical differences were observed when applying a test t-student between the presented approach and each of the state-of-the-art methods: Liu (p-value<0.05), Wang (p-value<0.05), and Kim (p-value <0.05).

Table 1: Mean (standard deviation) for each of the K-folds used to classify data.

Method	k	Accuracy	AUC	F-measure
STSalmap	3	**0.74 (0.08)**	**0.84 (0.04)**	**0.72 (0.08)**
	5	**0.72 (0.10)**	**0.82 (0.09)**	**0.71 (0.13)**
	7	**0.74 (0.11)**	**0.82 (0.08)**	**0.71 (0.15)**
Liu	3	0.51 (0.08)	0.55 (0.11)	0.47 (0.12)
	5	0.49 (0.11)	0.54 (0.15)	0.46 (0.14)
	7	0.42 (0.13)	0.43 (0.19)	0.40 (0.14)
Wang	3	0.51 (0.01)	0.50 (0.00)	-
	5	0.52 (0.02)	0.50 (0.00)	-
	7	0.52 (0.03)	0.50 (0.00)	-
Kim	3	0.72 (0.08)	0.80 (0.08)	0.72 (0.09)
	5	0.72 (0.12)	0.80 (0.12)	**0.71 (0.13)**
	7	0.72 (0.14)	0.80 (0.14)	**0.71 (0.15)**

Experiment 2: Training with a dataset and testing with the others. 20 control subjects and 20 HCM subjects from the ACDC dataset were used only for the training phase, while the testing dataset was composed of all subjects from MI09, YORK, and MI11 datasets (13 control and 15 HCM). Table 2 shows the accuracy, sensitivity, specificity, precision, AUC, and F-measure results obtained by the STSalmap descriptor and the baseline methods. Note how the STSalmap descriptor largely outperformed the baseline methods for all these metrics, reaching about 90% for both the F-measure and accuracy. In addition, the AUC of the Receiver Operating Characteristics (ROC) curve is shown for each method in Fig. 4. The AUC ROC curve of the STSalmap descriptor (0.94), plotted in blue, represents by far the best model. These results demonstrate these regional motion patterns may discriminate control and HCM subjects, with high precision and recall values (AUC of 94%), while baseline methods clearly failed at this task, i.e., AUC values lower than 71%

Table 2: Classification performance of the test group, 28 subjects (13 control subjects and 15 subjects with hypertrophic cardiomyopathy). Baseline methods: Liu et al. (2014), Wang et al. (2015) and Kim et al. (2015).

Measure	STSalmap	Liu	Wang	Kim
Accuracy	**0.89**	0.36	0.57	0.61
Sensitivity	**0.92**	0.38	0.54	0.62
Specificity	**0.87**	0.33	0.60	0.60
Precision	**0.86**	0.33	0.54	0.57
AUC	**0.94**	0.36	0.59	0.71
F-measure	**0.89**	0.36	0.54	0.59

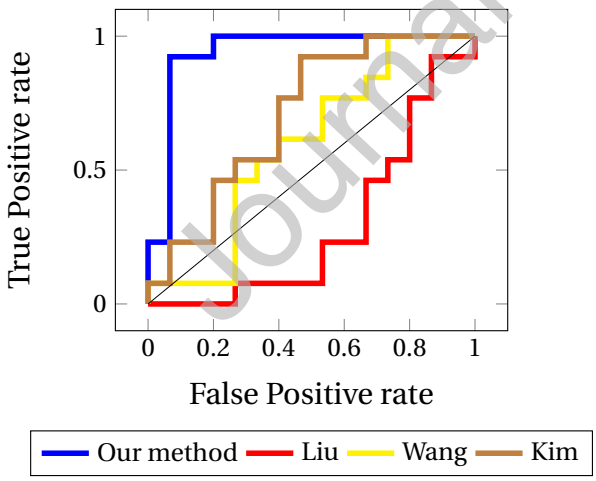


Fig. 4: ROC curve of the 28 testing. Blue, red, yellow and brown lines correspond to the STSalmap, Liu, Wang and Kim, respectively.

Figure 5 illustrates the short-axis view at basal and mid-levels (in rows) of four different subjects at the ED and their respective STSalmap (color map), superimposed to each of the corresponding slices. Saliency in these images highlights myocardium zones with higher motion in red and areas with little movement in blue. Illustration of these sections is organized, the control in the first two columns, and the HCM in third and fourth columns Fig. 5. These subjects are selected from different databases, demonstrating the method independence of the particular acquisition protocol or the anatomic cardiac level. As previously mentioned, saliency provides quantifiable heart motion information, and interestingly, the first two columns in Fig. 5 illustrate how control subjects show redder LV movement patterns, i.e, motion is more important and homogeneously observed for the whole ventricular structure along with the diastolic phase, in contrast to HCM subjects who appear to concentrate the ventricular movement in the lateral wall (red), while the septum and inferior segments show little motion (blue areas), a finding widely reported in the literature in case of HCM patients (Rickers et al. (2005); Mishiro et al. (1999)).

5.5. *Regional and local motion patterns for control and pathological subjects*

Figure 5 illustrates the short-axis view at basal and mid-levels (in rows) of four different subjects at the ED and their respective STSalmap (color map), superimposed to each of the corresponding slices. Significant differences, between control and HCM groups, were established by a t-test of the saliency dispersion ($\alpha = 0.05$, p-value < 0.05) for the LV segments. Two tests were here performed: the former compared the saliency dispersion at the four anatomical LV segments (septal, anterior, lateral, and inferior) of the whole ventricular volume, and the latter compared the local saliency dispersion at different ventricular levels, specifically, basal, mid and apical sections, aiming to show the aptness of the STSalmap to describe local motion patterns. For doing so, in this second test, a different ventricular partition was obtained by dividing in half the septal and lateral segments of the basal and mid sections. This partition was then composed of sixteen segments, the usual four of each of the three ventricular levels, and the new subdivisions at basal and mid-levels, that is to say, anterior-septal, inferior-septal, anterior-lateral, and inferior-lateral, following the AHA protocol (Cerqueira et al. (2002).)

5.5.1. *Regional motion patterns in control and HCM*

Saliency dispersion, defined in Eq. 4.3.2, was computed for each of four LV segments (septal, anterior, inferior, and lateral), at the ES and ED, for 20 control and 20 HCM subjects from the ACDC dataset. Saliency at each time was estimated by using only frames of each phase, and average and standard deviation were computed for each anatomical segment. Figure 6 illustrates the saliency dispersion average (lines) and standard deviation (shaded area) for control (in blue) and HCM (in red). Overall, control subjects exhibit a lower and more stable dispersion along the LV myocardium wall than HCM subjects, during both cycle phases. Significant differences of the dispersion values between control and HCM were found for septal (p-value < 0.05) and anterior (p-value < 0.05) segments at both the ES and ED. This higher dispersion in septal and anterior segments of HCM subjects might be attributed to the compensatory and dissimilar hypertrophy re-

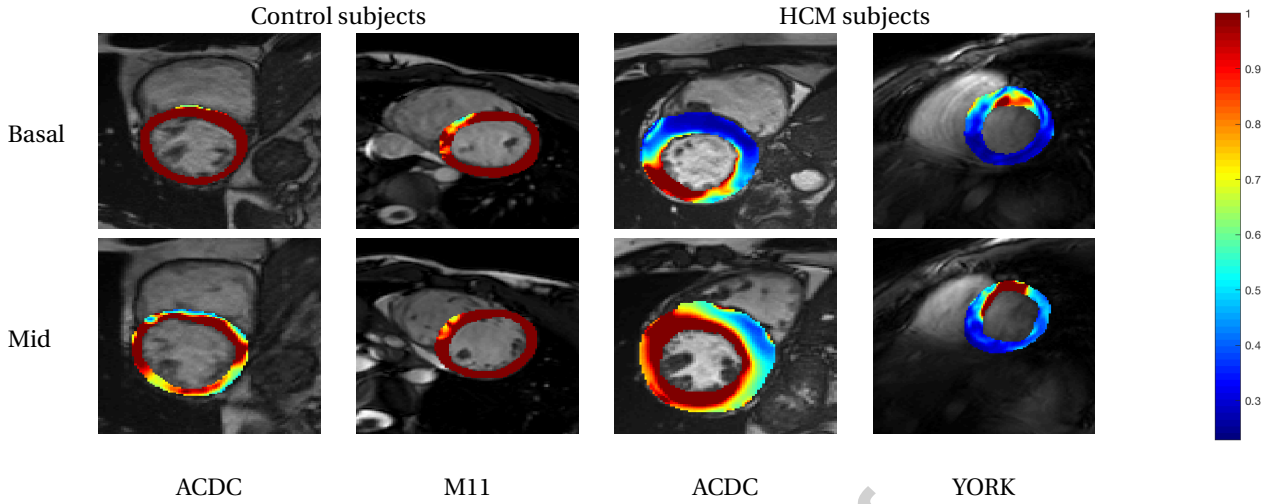


Fig. 5: Spatio-temporal saliency (color map) superimposed with the short axis cine MRI images of different subjects from different datasets at the ED.

action in these patients which produces complex motion patterns during diastolic and systolic phases. In contrast, control and HCM subjects exhibit more similar dispersion patterns at the inferior and lateral segments at both phases, yet differences were also found to be significant (p-value < 0.05), but in general smaller than those observed in the other two segments. Notice also how the dispersion trend for both phases, in the control group, shows a slight decrease at the anterior, inferior, and lateral segments, while the septal dispersion is conserved. This pattern is also observed in the HCM group, even though it shows higher values, but the septal dispersion increases at the ES. This particular value may suggest hypertrophy is more concentrated in septal and anterior areas, the reason why the contractility is usually impaired in these segments, a well-known sign of the hypertrophic cardiomyopathy (Rickers et al. (2005); Mishiro et al. (1999); Betocchi et al. (1993); Garg et al. (2017)).

5.5.2. *The saliency as a metrics of cardiac impairment*

Regional motion patterns are here mapped to a feature space with eight dimensions (saliency dispersion of the four segments at the ED and ES). This space is complex by nature and it is not straightforward how to compare multidimensional points and how motion information may be distributed. In such space, the Mahalanobis distance determines a particular grouping, control, and HCM, which may assume other forms than spherical. This distance was normalized by the maximum distance to provide a reliable comparison between both groups. Figure 7 illustrates a 2D projection of such space. For doing so, any subject (multidimensional point) is plotted as follows, the x coordinate corresponds to the Mahalanobis distance of the point to the control group and the y coordinate stands for the Mahalanobis distance to the HCM group. Therefore, the higher the distance to the control group, the higher the case probability of having HCM. Overall, HCM subjects (red triangles) are concentrated in an HCM zone (red area), where the distance to the HCM group is small and to the control group is variable. Likewise, control subjects were concentrated in the blue zone, where the distance to the control group is small. Only one control subject obtained a small distance to the HCM group. This subject shows little deformation in the septal segment when visually inspected. Interestingly, this analysis locates individuals with a high probability of belonging to the correct class. This analysis quantifies the degree of motion impairment and eventually could be used as a potential risk indicator of HCM.

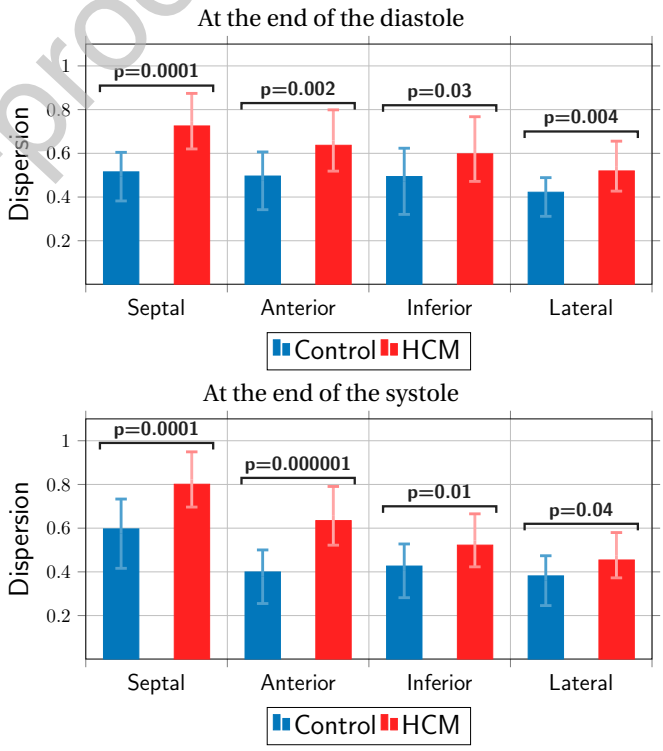


Fig. 6: Mean and standard deviation of dispersion values for each of four segments of the LV ventricular wall and the control (in blue) and HCM (in red) subjects from the ACDC dataset at the ED and ES.

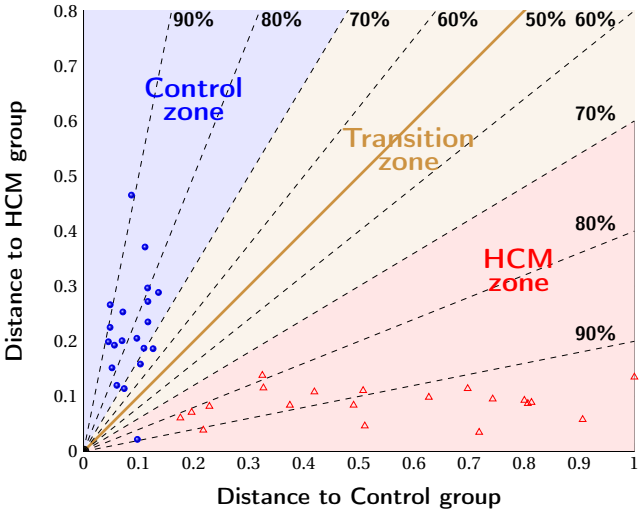


Fig. 7: 2D projection of the feature space, x coordinate corresponds to the Mahalanobis distance of any multidimensional point to the control group while the y coordinate represents the Mahalanobis distance to the HCM group. The golden line stands for the division between HCM and control subjects (provided in the dataset). Transition zone represents a classification error probability of about 30% for each group.

5.5.3. *Localized motion patterns in control, HCM, DCM and MINF*

A more local analysis was also performed by dividing the LV into sections (basal, mid, and apical) and the LV wall in anatomical segments (septal, inferior, lateral, and anterior). In basal and mid sections, the septal segment was divided into anteroseptal and inferoseptal, and the lateral segment was divided into anterolateral and inferolateral. All these divisions were automatically obtained as described in Subsection 4.3.1 by following the AHA model (Cerqueira et al. (2002)). The saliency dispersion was computed for each of the anatomical segments and sections at the ED and ES for each of 20 control, 20 HCM, 20 DCM, and 20 MINF subjects from the ACDC dataset. The average of the saliency dispersion for each segment and section at each time was computed to illustrate a global pattern (Fig. 8). Note how control subjects show a more homogeneous motion along the LV wall at the ED. At the end of the systole, a significant difference between control and MINF groups was found for the mid/apical anterior segments (test t-student, p-value < 0.05) and mid/apical inferior segments (test t-student, p-value < 0.05). These differences correlate with the impaired rotation in the LV apical/basal sections reported in MINF subjects (Park et al. (2011)). DCM also showed significant differences with control subjects at the ES in the mid/apical anterior segments (test t-student, p-value < 0.05) and the basal/mid-inferior and all apical segments (test t-student, p-value < 0.05), results that might be related with impaired apical rotation, as reported in DCM (Kanzaki et al. (2006)). As visualized in Fig. 6, control and HCM subjects presented a slight homogeneous motion in the lateral wall at both phases. While the septal and anterior segments in HCM evidenced a mix of movements. At the ED, these differences were also noted, with slight differences concerning the ES.

Note how multiple t-tests were performed per segment, and separately per cardiac phase: control group vs MINF group, control group vs HCM group, and control group vs DCM group. We obtained 48 t-tests per cardiac phase, at the ED and ES. Since p-value <0.05 was established in the conclusions, finding significant differences between groups per segment for each cardiac phase, we computed the Bonferroni correction to see the significance levels corrected for multiple testing. For doing so, the Bonferroni correction was estimated as 0.05/48=0.001. At the ES, we found significant differences in all segments except for 1st, 2nd, 5th, 6th, 9th, and 14th (anterior and lateral segments in the basal region), in which similar motion patterns were observed in all groups. While at the ED, the DCM group obtained no significant difference in the 4th and 5th segments. HYP obtained similar motion patterns in segments 5th, 6th, 10th, 11th, 12th, and 14th segments (infero and antero lateral segments in basal section and apical region). Interestingly, no significant differences were reported for the first segments in the MINF group: 2nd, 3rd, 5th, 6th, 8th, 9th, and 10th, i.e., in all segments from the basal region.

These findings show the STSalmap descriptor can extract relevant information of the systolic and diastolic function (movement patterns) at any measurable location. It can be correlated with pathological findings. This analysis can be carried out per patient to detect local abnormal movement.

5.6. *Method performance on noisy images*

In addition, the presented saliency descriptor was robust to noise when Rician, speckle, and Gaussian noise were separately added to the same cases classified in experiment 2 of Section 5.4. Specifically, speckle noise with mean 0 and variance 0.05 was added to the images, while the Rician noise was configured with mean 0 and 5 Gaussian noise levels. These images were re-normalized by using the min-max algorithm. Results using the two first types of noise did not exhibit differences with the original images (data not shown). A more exhaustive analysis was performed for the Gaussian noise, a more typical MRI noise, by testing different parameter configurations (mean and standard deviation). This analysis is observed in Table 3. The SVM model was trained with the ACDC dataset, contaminated with the Gaussian noise, and tested in the other four datasets. Observe how the descriptor maintains comparable accuracy when the mean and standard deviation are set to 0.3 and relative deterioration is observed when increasing these values, the descriptor discriminates up to 50% of the data in a very noisy environment.

5.7. *Time-performance analysis*

All experiments were performed in a computer with 2.9 GHz CPU and 16 GB of RAM, using a simple Matlab implementation with no training. The average of the volume dimensions was of $241.8 \pm 25.3 \times 255.2 \pm 19.4 \times 10.8 \pm 2.5$. The estimated STSalmap saliency was computed in 3D, unlike the state-of-the-art methods which compute this in 2D. The time in seconds (s) to obtain a salient volume by the

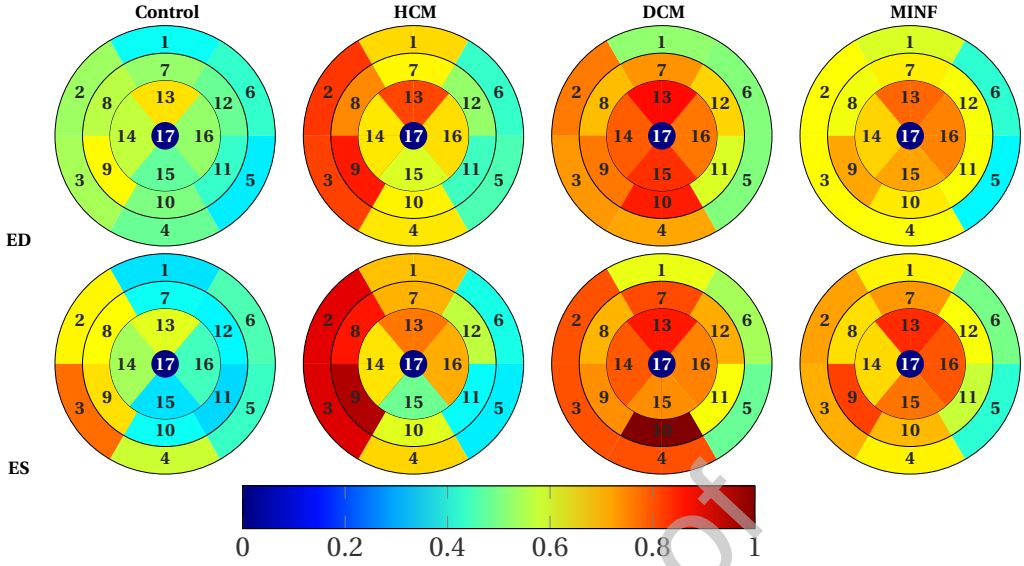


Fig. 8: Bull's eye plot Bakos et al. (2014) of the saliency dispersion average in different anatomical segments and sections, at the ED and ES (ACDC dataset). LV segments were obtained following the AHA protocol (Cerqueira et al. (2002)). Segments one to six are: anterior, anteroseptal, inferoseptal, inferior, inferolateral and anterolateral at the basal section. Segments seven to twelve correspond to the same previous segments but at the mid section while segments thirteen to sixteen are the anterior, septal, inferior, and lateral at the apical section. The 17^{th} segment is the apex, but this information was not included. Four groups are compared: Control, hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), and myocardial infarction (MINF).

Table 3: Accuracy for each Gaussian noise configuration (mean and standard deviation) added to data for the classification task.

	standard deviation			
mean	**0.3**	**0.5**	**0.7**	**1**
0.3	0.85	0.65	0.68	0.65
0.5	0.73	0.55	0.58	0.50
0.7	0.53	0.50	0.60	0.68
1	0.68	0.63	0.63	0.70

STSalmap descriptor was about 12.12 *s*, while obtained times from other methods to compute all the 2D slices of a whole volume were 51.31 *s* Liu et al. (2014), 45.30 *s* Wang et al. (2015) and 81.93 *s* Kim et al. (2015). The saliency descriptor code could be available by contacting the corresponding author directly and under the conditions of citing this article in any scientific publication and, it will be published in https://github.com/amatehortual/MSMS.

6. Discussion

This work introduced a novel spatio-temporal saliency descriptor to characterize and quantify motion cardiac patterns independently of the acquisition protocol. This saliency descriptor describes and summarizes relevant temporal information of the cardiac function, no matter the particular angle or resolution of the captured video. The underlying idea of this model is to mimic the visual motion perception and extract relevant features by decomposing the scene into regions at multiple scales and detecting local-oriented motion associated with spatio-temporal changes. After these estimations are integrated at the multiple orientations and scales, a coherent regional motion sets a local saliency. This descriptor highlights a relative motion history of different structures which is correlated with normal and abnormal cardiac patterns. This work presents, in addition, a statistical analysis of the ventricular differences in different cardiac entities. Finally, this descriptor was applied to cine MRI to associate local-regional motion with control and pathological patterns, correlating them in terms of well-known anatomic findings.

Several works have characterized myocardial motion patterns to discriminate controls and pathological subjects in cine MRI from homogenous datasets. Among them, techniques such as optical flow, free-form deformation, block matching, and deep learning were used to extract velocity features Peressutti et al. (2017); Puyol-Antón et al. (2017); Puyol-Antón et al. (2019); Bello et al. (2019); Piras et al. (2017); Morales et al. (2019); Terpstra et al. (2020). Other methods segmented the ventricles in the mid-slice of cine MRI to track the geometric features (thickness, radius) over time Zheng et al. (2019). However, these related works present some disadvantages. These methods are prone the error yet, the noise accentuates the particular pathology since adaptation mechanisms increase the variability of motion to both, the cardiac and nearby thoracic structures. In another way, impaired apical and basal rotations, specific to myocardial infarction and diastolic cardiomyopathy (Park et al. (2011); Kanzaki et al. (2006)), would be impossible to characterize with these methods since the analysis was focused in the mid-slice of cine

MRI. These methods require correcting large affine and non-linear transformations associated with respiratory motion. These works in general require still to prove their aptness to generalize their results in heterogeneous databases and to cope with the large range of cardiac anatomic deformations and the particular acquisition protocols.

Unlike the mentioned approaches, the presented spatio-temporal saliency captures motion trends and, as herein demonstrated, they can be associated with several pathological patterns. The estimated saliency stands for a quantified motion, i.e., slow and rapid movements, evidence of the myocardial stage when contracting, relaxing, twisting, untwisting, lengthening, and shortening over time. Because some segments of the myocardium are more affected by the disease, the saliency dispersion was computed for each of the standardized left ventricular segments. A high inter-and-intra saliency variability in the segments reflected an abnormal motion pattern, well correlated with disease physiopathology, for instance, the reduced apex motion in patients with infarction. The proposed descriptor captured non-homogeneous movement in the septal and anterior regions, at the end of the diastole and the end of the systole, signs of hypertrophic cardiomyopathy (Marian and Braunwald (2017)). Likewise, the high dispersion visualized in the apical-anterior, mid and apical-inferior segments in subjects with dilated cardiomyopathy are well correlated with the impaired torsional dynamics of the left ventricle, a finding widely acknowledged as a marker of the disease severity (Popescu et al. (2009)). These findings evidence that saliency can quantify systolic and diastolic dysfunction at local and regional levels, a definite method plus in terms of clinical interpretability. Moreover, the presented descriptor demonstrated to be generalizable since it dealt with several heterogeneous datasets. Furthermore, only two parameters required adjustments, namely the patch size ($7 \times 7 \times 7$) and σ (herein set to 3), an attractive matter of this method because it decreases the dependence on data, as it may be the case of other approaches like deep learning, whose robustness is limited by the distribution and size of the training set.

Several limitations of the presented method may be: i) The sum for the set of orientations from 0° to 315° might remove the main motion direction, which could be useful in motion-related tasks. However, the herein obtained saliency estimates a quantity about the motion history and the relative temporal homogeneity within a region along the time. This characterization is well-correlated with known described disease findings. ii) Comparison with regional cardiac motion descriptors such as regional displacement, velocity, strain, or was not possible since this information was not provided by the available dataset. However, strain and motion tracking methods could also be used for identifying wall motion abnormalities Puyol-Antón et al. (2018); Vo et al. (2018). iii) Comparison with state-of-the-art motion saliency methods can be considered unfair since they were not optimally parameterized. However, their source codes were closed and only executable codes were available. iv) Motion patterns characterized by the proposed method can potentially provide new insights of the physio-pathology to devise risk indicators adapted to the myocardial dysfunction. The fact of following up any instant of the cardiac cycle is an advantageous side effect that can be used to design local/global indicators of the temporal cardiac function. disease severity (Popescu et al. (2009)). These findings evidence that saliency can quantify systolic and diastolic dysfunction at local and regional levels, a definite method plus in terms of clinical interpretability. Moreover, the presented descriptor demonstrated to be generalizable since it dealt with several heterogeneous datasets. Furthermore, only two parameters required adjustments, namely the patch size ($7 \times 7 \times 7$) and σ (herein set to 3), an attractive matter of this method because it decreases the dependence on data, as it may be the case of other approaches like deep learning, whose robustness is limited by the distribution and size of the training set.

Despite these limitations, results evidenced that the proposed framework was able to quantify regional motion patterns associated with cardiac function, an impossible task to reach just by visual inspection of the data. The proposed framework may potentially help to detect and quantify subtle changes in the myocardial function at any time of the cardiac cycle. This study may complement current strategies of cardiac function assessment and can help to find out myocardial dysfunction with more individualized indicators of the cardiac output Weissler-Snir et al. (2017). Likewise, the proposed investigation can be applied to both right and left ventricles, achieving a saliency of the whole sequence. An interventricular analysis may provide a description more detailed of cardiac dynamics, and a quantitative progression of heart disease.

7. Conclusion

The presented method demonstrated robustness and stability at extracting relevant information in short axis cine-MRI images from heterogeneous datasets with images acquired from different equipment and diseases. This saliency information is demonstrated to be useful at highlighting and quantifying abnormal movement patterns during the diastolic and systolic functions associated with hypertrophic cardiomyopathy. Results demonstrate this method correlates with abnormal patterns of the disease at local and regional cardiac levels. In addition, several medical applications could benefit from this descriptor to improve disease understanding (Jan and Tajik (2017)), specifically assessing myocardium function, extracting information for gait and eye motion analysis, and designing adapted indicators for risk management of cardiac or brain complications in angiography, among others. Future research will focus on the characterization of cardiac pathologies in a larger dataset of cine MRI images.

Conflicts of Interest

None.

Acknowledgments

This work was supported by Colciencias-Colombia, Grant No. 647 (2015 call for National PhD studies), and Région Bretagne in the framework of the Investissement d'Avenir Program through Labex CAMI (ANR-11-LABX-0004).

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. Slic superpixels compared to state-of-the-art superpixel methods. IEEE transactions on pattern analysis and machine intelligence 34, 2274–2282.

Adelson, E., Bergen, J., 1985. Spatiotemporal energy models for the perception of motion. Josa a 2, 284–299.

Albright, T.D., Stoner, G.R., 1995. Visual motion perception. Proceedings of the National Academy of Sciences 92, 2433–2440.

Allard, R., Arleo, A., 2017. Reducing luminance intensity can improve motion perception in noise. Scientific reports 7, 43140–43144.

Andreopoulos, A., Tsotsos, J., 2008. Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI. Medical Image Analysis 12, 335–357.

Bak, C., Kocak, A., Erdem, E., Erdem, A., 2018. Spatio-temporal saliency networks for dynamic saliency prediction. IEEE Transactions on Multimedia 20, 1688–1698.

Bakos, Z., Markstad, H., Ostenfeld, E., Carlsson, M., Roijer, A., Borgquist, R., 2014. Combined preoperative information using a bullseye plot from speckle tracking echocardiography, cardiac ct scan, and mri scan: targeted left ventricular lead implantation in patients receiving cardiac resynchronization therapy. European Heart Journal–Cardiovascular Imaging 15, 523–531.

Bello, G.A., Dawes, T.J.W., Duan, J., Biffi, C., de Marvao, A., Howard, L.S.G.E., et al., 2019. Deep-learning cardiac motion analysis for human survival prediction. Nature machine intelligence 1, 95–104.

Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P., et al., 2018. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? IEEE transactions on medical imaging 37, 2514–2525.

Betocchi, S., Hess, O., Losi, M.A., Nonogi, H., Krayenbuehl, H., 1993. Regional left ventricular mechanics in hypertrophic cardiomyopathy. Circulation 88, 2206–2214.

Bradley, D., Goyal, M., 2008. Velocity computation in the primate visual system. Nature Reviews Neuroscience 9, 686.

Buckberg, G., Hoffman, J.I., Mahajan, A., Saleh, S., Coghlan, C., 2008. Cardiac mechanics revisited. Circulation 118, 2571–2587.

Burt, P., Adelson, E., 1983. The laplacian pyramid as a compact image code. IEEE Transactions on Communications 31, 532–540.

Burt, P., Adelson, E., 1987. The laplacian pyramid as a compact image code, in: Readings in Computer Vision. Elsevier, pp. 671–679.

Cerqueira, M.D., Weissman, N.J., Dilsizian, V., Jacobs, A.K., Kaul, S., Laskey, W.K., et al., 2002. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: a statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the american heart association. Circulation 105, 539–542.

Chen, Y., Zou, W., Tang, Y., Li, X., Xu, C., Komodakis, N., 2018. Scom: Spatiotemporal constrained optimization for salient object detection. IEEE Transactions on Image Processing 27, 3345–3357.

Ding, Y., Sohn, J.H., Kawczynski, M.G., Trivedi, H., Harnish, R., Jenkins, N.W., et al., 2019. A deep learning model to predict a diagnosis of alzheimer disease by using 18f-fdg pet of the brain. Radiology 290, 456–464.

Dubost, F., Adams, H., Bortsova, G., Ikram, M.A., Niessen, W., Vernooij, M., [de Bruijne], M., 2019. 3d regression neural network for the quantification of enlarged perivascular spaces in brain mri. Medical Image Analysis 51, 89 – 100.

Editorial, 2018. Taking personalized medicine to heart. Nature Medicine 24, 113.

Elliott, P.M., Anastasakis, A., Borger, M.A., Borggrefe, M., Cecchi, F., Charron, P., et al., 2014. 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: The Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC). European Heart Journal 35, 2733–2779.

Faisal, M., Akhter, I., Ali, M., Hartley, R., 2020. Epo-net: Exploiting geometric constraints on dense trajectories for motion saliency, in: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE. pp. 1873–1882.

Florack, L., ter Haar Romeny, B., Koenderink, J., Viergever, M., 1992. Scale and the differential structure of images. Image and vision computing 10, 376–388.

Freeman, W., Adelson, E., 1991. The design and use of steerable filters. IEEE Transactions on Pattern Analysis & Machine Intelligence , 891–906.

Garg, S., de Lemos, J.A., Matulevicius, S.A., Ayers, C., Pandey, A., Neeland, I.J., et al., 2017. Association of concentric left ventricular hypertrophy with subsequent change in left ventricular end-diastolic volume: the dallas heart study. Circulation: Heart Failure 10, 3959–3966.

Huang, C.L., Chen, Y.T., 1995. Motion estimation method using a 3d steerable filter. Image and Vision Computing 13, 21–32.

Huang, R., Xie, W., Noble, A., 2018. Vp-nets: Efficient automatic localization of key brain structures in 3 D fetal neurosonography. Medical image analysis 47, 127–139.

Iakovidis, D., Georgakopoulos, S., Vasilakakis, M., Koulaouzidis, A., Plagianakos, V., 2018. Detecting and locating gastrointestinal anomalies using deep learning and iterative cluster unification. IEEE Transactions on Medical Imaging 37, 2196–2210.

Jan, F., Tajik, J., 2017. Modern imaging techniques in cardiomyopathies. Circulation Research 121, 874–891.

Jiang, L., Xu, M., Liu, T., Qiao, M., Wang, Z., 2018. Deepvs: A deep learning based video saliency prediction approach, in: Computer Vision–ECCV 2018. Springer, pp. 625–642.

Johansson, G., 1973. Visual perception of biological motion and a model for its analysis. Perception & psychophysics 14, 201–211.

Kanzaki, H., Nakatani, S., Yamada, N., Urayama, S.i., Miyatake, K., Kitakaze, M., 2006. Impaired systolic torsion in dilated cardiomyopathy: reversal of apical rotation at mid-systole characterized with magnetic resonance tagging method. Basic research in cardiology 101, 465–470.

Kervrann, C., Boulanger, J., 2006. Optimal spatial adaptation for patch-based image denoising. IEEE Transactions on Image Processing 15, 2866–2878.

Kim, H., Kim, Y., Sim, J.Y., Kim, C.S., 2015. Spatiotemporal saliency detection for video sequences based on random walk with restart. IEEE Transactions on Image Processing 24, 2552–2564.

Kramer, C.M., Reichek, N., Ferrari, V.A., Theobald, T., Dawson, J., Axel, L., 1994. Regional heterogeneity of function in hypertrophic cardiomyopathy. Circulation 90, 186–194.

Krittanawong, C., Johnson, K.W., Rosenson, R.S., Wang, Z., Aydar, M., Baber, U., et al., 2019. Deep learning for cardiovascular medicine: a practical primer. European heart journal 40, 2058–2073.

Leiner, T., Rueckert, D., Suinesiaputra, A., Baeßler, B., Nezafat, R., Išgum, I., Young, A.A., . Machine learning in cardiovascular magnetic resonance: basic concepts and applications. Journal of Cardiovascular Magnetic Resonance 21, 61–74.

Liu, Z., Li, J., Ye, L., Sun, G., Shen, L., 2017. Saliency detection for unconstrained videos using superpixel-level graph and spatiotemporal propagation. IEEE transactions on circuits and systems for video technology 27, 2527–2542.

Liu, Z., Zhang, X., Luo, S., Le Meur, O., 2014. Superpixel-based spatiotemporal saliency detection. IEEE transactions on circuits and systems for video technology 24, 1522–1540.

Maczyta, L., Bouthemy, P., Le Meur, O., 2019. Cnn-based temporal detection of motion saliency in videos. Pattern Recognition Letters 128, 298–305.

Mahapatra, D., Buhmann, J., 2016. Visual saliency-based active learning for prostate magnetic resonance imaging segmentation. Journal of Medical Imaging 3, 14003–14013.

Mann, D.L., Bristow, M.R., 2005. Mechanisms and models in heart failure: the biomechanical model and beyond. Circulation 111, 2837–2849.

Marian, A.J., Braunwald, E., 2017. Hypertrophic cardiomyopathy: genetics, pathogenesis, clinical manifestations, diagnosis, and therapy. Circulation research 121, 749–770.

Massanes, F., Brankov, J.G., 2011. Motion perception in medical imaging, in: Medical Imaging 2011: Image Perception, Observer Performance, and Technology Assessment, International Society for Optics and Photonics. pp. 796610–796619.

Mather, G., Pavan, A., Bellacosa Marotti, R., Campana, G., Casco, C., 2013. Interactions between motion and form processing in the human visual system. Frontiers in computational neuroscience 7, 65.

Mazein, A., Ostaszewski, M., Kuperstein, I., Watterson, S., Le Novère, N., Lefaudeux, D., et al., 2018. Systems medicine disease maps: community-driven comprehensive representation of disease mechanisms. NPJ systems biology and applications 4, 21–30.

Mishiro, Y., Oki, T., Iuchi, A., Tabata, T., Yamada, H., Abe, M., et al., 1999. Regional left ventricular myocardial contraction abnormalities and asynchrony in patients with hypertrophic cardiomyopathy evaluated by magnetic resonance spatial modulation of magnetization myocardial tagging. Japanese circulation journal 63, 442–446.

Morales, M.A., Izquierdo-Garcia, D., Aganj, I., Kalpathy-Cramer, J., Rosen, B.R., Catana, C., 2019. Implementation and validation of a three-dimensional cardiac motion estimation network. Radiology: Artificial Intelligence 1, e180080.

Nguyen, T., Wu, Q., 2016. Multiple kernel point set registration. IEEE Transactions on Medical Imaging 35, 1381–1394.

Oksuz, I., Ruijsink, B., Puyol-Antón, E., Clough, J.R., Cruz, G., Bustin, A., et al., 2019. Automatic cnn-based detection of cardiac MR motion artefacts using k-space data augmentation and curriculum learning. Medical image analysis 55, 136–147.

Omar, A.M.S., Vallabhajosyula, S., Sengupta, P.P., 2015. Left ventricular twist and torsion. Circulation: Cardiovascular Imaging 8, 3029–3035.

Park, S.M., Hong, S.J., Ahn, C.M., Kim, Y.H., Kim, J.S., Park, J.H., et al., 2011. Different impacts of acute myocardial infarction on left ventricular apical and basal rotation. European Heart Journal - Cardiovascular Imaging 13, 483–489.

Patil, P.W., Thawakar, O., Dudhane, A., Murala, S., 2019. Motion saliency based generative adversarial network for underwater moving object segmentation, in: 2019 IEEE International Conference on Image Processing (ICIP), IEEE. pp. 1565–1569.

Peressutti, D., Sinclair, M., Bai, W., Jackson, T., Ruijsink, J., Nordsletten, D., et al., 2017. A framework for combining a motion atlas with non-motion information to learn clinically useful biomarkers: Application to cardiac resynchronisation therapy response prediction. Medical Image Analysis 35, 669 – 684.

Piras, P., Teresi, L., Puddu, P.E., Torromeo, C., Young, A.A., Suinesiaputra, A., Medrano-Gracia, P., 2017. Morphologically normalized left ventricular motion indicators from mri feature tracking characterize myocardial infarction. Scientific reports 7, 1–14.

Popescu, B.A., Beladan, C.C., Călin, A., Muraru, D., Deleanu, D., Roşca, M., Ginghină, C., 2009. Left ventricular remodelling and torsional dynamics in dilated cardiomyopathy: reversed apical rotation as a marker of disease severity. European journal of heart failure 11, 945–951.

Puyol-Antón, E., Ruijsink, B., Bai, W., Langet, H., De Craene, M., Schnabel, J.A., Piro, P., King, A.P., Sinclair, M., 2018. Fully automated myocardial strain estimation from cine mri using convolutional neural networks, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE. pp. 1139–1143.

Puyol-Antón, E., Ruijsink, B., Gerber, B., Amzulescu, M.S., Langet, H., De Craene, M., et al., 2019. Regional multi-view learning for cardiac motion analysis: Application to identification of dilated cardiomyopathy patients. IEEE Transactions on Biomedical Engineering 66, 956–966.

Puyol-Antón, E., Sinclair, M., Gerber, B., Amzulescu, M.S., Langet, H., Craene, M.D., et al., 2017. A multimodal spatiotemporal cardiac motion atlas from mr and ultrasound data. Medical Image Analysis 40, 96 – 110.

Radau, P., Lu, Y., Connelly, K., Paul, G., Dick, A., Wright, G., 2009. Evaluation framework for algorithms segmenting short axis cardiac MRI. The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge 49.

Rickers, C., Wilke, N.M., Jerosch-Herold, M., Casey, S.A., Panse, P., Panse, N., et al., 2005. Utility of cardiac magnetic resonance imaging in the diagnosis of hypertrophic cardiomyopathy. Circulation 112, 855–861.

Sanchez-Martinez, S., Duchateau, N., Erdei, T., Aakhus, S., Degiovanni, A., Marino, P., et al., 2018. Machine learning analysis of left ventricular function to characterize heart failure with preserved ejection fraction. Circulation: Cardiovascular Imaging 11, 7138–7149.

Sarmiento, E., Pico, J., Martinez, F., 2018. Cardiac disease prediction from spatio-temporal motion patterns in cine-mri, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE. pp. 1305–1308.

Schrater, P., Knill, D., Simoncelli, E., 2000. Mechanisms of visual motion detection. Nature neuroscience 3, 64.

Sun, M., Zhou, Z., Hu, Q., Wang, Z., Jiang, J., 2018. Sg-fcn: A motion and memory-based deep learning model for video saliency detection. IEEE transactions on cybernetics 49, 2900–2911.

Tang, Y., Zou, W., Jin, Z., Li, X., 2018. Multi-scale spatiotemporal conv-lstm network for video saliency detection, in: Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval, ACM. pp. 362–369.

Terpstra, M.L., Maspero, M., D'Agata, F., Stemkens, B., Intven, M.P., Lagendijk, J.J., et al., 2020. Deep learning-based image reconstruction and motion estimation from undersampled radial k-space for real-time mri-guided radiotherapy. Physics in Medicine & Biology , 2–24.

Tobon-Gomez, C., Craene], M.D., McLeod, K., Tautz, L., Shi, W., Hennemuth, A., et al., 2013. Benchmarking framework for myocardial tracking and deformation algorithms: An open access database. Medical image analysis 17, 632–648.

Treue, S., Maunsell, J., 1996. Attentional modulation of visual motion processing in cortical areas mt and mst. Nature 382, 539.

Vo, H.Q., Marwick, T.H., Negishi, K., 2018. Mri-derived myocardial strain measures in normal subjects. JACC: Cardiovascular Imaging 11, 196–205.

Wang, W., Shen, J., Shao, L., 2015. Consistent video saliency using local gradient flow optimization and global refinement. IEEE Transactions on Image Processing 24, 4185–4196.

Wang, W., Shen, J., Yang, R., Porikli, F., 2018. Saliency-aware video object segmentation. IEEE Transactions on Pattern Analysis & Machine Intelligence , 20–33.

Weissler-Snir, A., Adler, A., Williams, L., Gruner, C., Rakowski, H., 2017. Prevention of sudden death in hypertrophic cardiomyopathy: bridging the gaps in knowledge. European heart journal 38, 1728–1737.

Zhang, J., Saha, A., Zhu, Z., Mazurowski, M.A., 2018. Hierarchical convolutional neural networks for segmentation of breast tumors in mri with application to radiogenomics. IEEE transactions on medical imaging 38, 435–447.

Zhang, J., Xie, Y., Xia, Y., Shen, C., 2019. Attention residual learning for skin lesion classification. IEEE transactions on medical imaging 38, 2092–2103.

Zhao, Y., Zheng, Y., Liu, Y., Yang, J., Zhao, Y., Chen, D., et al., 2017. Intensity and compactness enabled saliency estimation for leakage detection in diabetic and malarial retinopathy. IEEE Transactions on Medical Imaging 36, 51–63.

Zheng, Q., Delingette, H., Ayache, N., 2019. Explainable cardiac pathology classification on cine mri with motion characterization by semi-supervised learning of apparent flow. Medical image analysis 56, 80–95.